AD______________

ARMY PROJECT ORDER NO: 88PP8805

TITLE: MICROSAINT APPLICATIONS: DOCUMENTATION OF JWGD3 MILPERF PERFORMANCE ASSESSMENT SYSTEM

PRINCIPAL INVESTIGATOR: Ronald P. Crisman, Ph.D.

CONTRACTING ORGANIZATION: Naval Aerospace Medical Research Laboratory, Research Department
Naval Air Station
Pensacola, Florida 32508-5700

REPORT DATE: August 30, 1990

TYPE OF REPORT: Final Report

PREPARED FOR: U.S. ARMY MEDICAL RESEARCH AND DEVELOPMENT COMMAND
Fort Detrick, Frederick, Maryland 21702-5012

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Naval Aerospace Medical Research Laboratory, Research Department 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Naval Air Station Pensacola, Florida 32508-5700	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U.S. Army Medical Research & Development Command 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER Army Project Order No. 88PP8805
8c. ADDRESS (City, State, and ZIP Code) Fort Detrick Frederick, Maryland 21702-5012	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 63807A PROJECT NO. 3M4 63807D993 TASK NO. BO WORK UNIT ACCESSION NO. 046

11. TITLE (Include Security Classification)
MICROSAINT APPLICATIONS: DOCUMENTATION OF JWGD3 MILPERF PERFORMANCE ASSESSMENT SYSTEM

12. PERSONAL AUTHOR(S)
Ronald P. Crisman, Ph.D.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final Report	FROM 1/1/88 TO 12/31/89	1990 August 30	14

16. SUPPLEMENTARY NOTATION
— fr 8. 1

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Human Performance; Modeling, Performance Assessment; Task Analysis; Volunteers; ~~RA V~~ (SIU) T
23	02		
05	05		

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

To make detailed and accurate predictions of the effects of chemical defense drugs on operational performance will require (1) increasing the detail in which both laboratory and operational performance measures are described and quantified and (2) improving the techniques used to generalize changes observed in the laboratory to tasks performed in operational settings. This final report outlines an approach to developing computational performance models that reflect the structure of laboratory and operational performance. Such models can substantially increase the detail in which performance is quantified. To effectively use such models in prediction will require developing procedures for linking the parameters of laboratory tasks to those of operational tasks. The linking process is discussed, along with some alternative approaches to prediction. Keywords.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Mrs. Virginia M. Miller	22b. TELEPHONE (Include Area Code) (301) 663-7325	22c. OFFICE SYMBOL SGRD-RMI-S

DD Form 1473, JUN 86 Previous editions are obsolete. SECURITY CLASSIFICATION OF THIS PAGE

FOREWORD

Opinions, interpretations, conclusions, and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

______ Where copyrighted material is quoted, permission has been obtained to use such material.

______ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

RS Citations of commercial organizations and trade names in this report do not constitute an official Department of the Army endorsement or approval of the products or services of these organizations.

______ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Animal Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

RS For the protection of human subjects, the investigator(s) have adhered to policies of applicable Federal Law 45CFR46.

______ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institute of Health.

31 AUG 90

Principal Investigator's Signature Date

CONTENTS

Page

FOREWORD iii

1. INTRODUCTION 1

2. METHODS 3

EMPIRICAL PERFORMANCE DATA 3

SIMULATION PROCEDURES 3

3. RESULTS AND DISCUSSION 4

4. CONCLUSIONS 6

REFERENCES 8

1. INTRODUCTION

This Final Report covers work performed at the Naval Aerospace Medical Research Laboratory during Fiscal Years 1988 and 1989. The Joint Working Group on Drug Dependent Degradation in Military Performance (JWGD3 MILPERF) was established for the purpose of developing and testing procedures to evaluate the effects of chemical defense pharmaceutical agents on military performance. The products of the JWGD3 have included tests, test batteries, task analysis systems, performance modeling tools, simulators, databases, and archives of human performance data. These tools, although specifically designed for chemical defense analyses, have been used to measure the effects of various interventions (or stressors) on military performance. Examples of such interventions and stressors are pharmaceuticals (including prophylactics, treatment drugs, and performance enhancing drugs), (e.g., sleep loss and acceleration), and environmental stressors (e.g., extremes of temperature; reference 1). to 1473

An objective of this laboratory's participation has been to develop computational models of human performance in operational tasks and in laboratory performance tests. The purpose has been to develop procedures that might be used to generalize laboratory measurements of human performance, such as those derived from the Unified Tri-services Cognitive Performance Test Battery (UTCPAB; 2-3), that would allow users to transform data from performance tests into detailed predictions about performance in operational systems. Such predictions might be performed by first analyzing the temporal organization of performance in a target operational system into elements and using these elements to build a model of the system. Test information might then be transferred between performance and operational models when an element is common to both (and when the information processing requirements and other contingencies of the system and the performance test are similar). The simplest example of such a transfer would occur when a parameter of an operational model element is set equal to its value in the corresponding test model. Dynamic examples would occur when an operational model parameter is caused to track changes in the corresponding test model parameter that occur as functions of other variables, such as time.

Work originally planned for Fiscal Year 1988 included developing a task-analytic model of performance in a helicopter simulator. This model was to have been merged with subsidiary models of the biological effects of antihistamines and used to predict the effects of antihistamines on performance in helicopter and (in a second effort) naval-tactical flight simulators. The work was originally to have been a collaborative effort involving at least three research projects from two different laboratories. Various factors combined to render that work unsuccessful. Two important contributing factors were personnel reassignments and difficulties encountered in meshing the logistics, instrumentation, and milestone schedules of the different projects. In Fiscal Year 1989, we focused the project on the narrower topic of developing techniques for modeling laboratory tests of human performance (see references 4-6). This allowed us to examine more adequately some questions regarding how performance test data might actually be integrated into models of operational tasks.[1]

The performance test models we have developed are driven by equations derived from empirical data. They were written in MicroSAINT, which is a task-simulation language that runs on personal computers. MicroSAINT is derived from the System Analysis of Integrated Networks of Tasks (SAINT, refer-

---

[1]We have proposed to revisit the issue of modeling flight performance in a separate project beginning in Fiscal Year 1991.

ence 7). SAINT is a computer-simulation language that runs on mainframe computers; it was developed for writing network performance models of the type introduced in human engineering during the 1960s by Siegel and Wolf (8).

Many of the performance models used in human engineering today appear to be derived from the Siegel-Wolf network approach. Models of this type differ substantially from the traditional control-theoretic and optimal-control models of human engineering. Control-theoretic models have typically used closed-loop stability analysis to generate functions describing the performance of man-machine system operators. The tasks most frequently addressed by such models are continuous, manual-control tasks. Optimal control models represent the performance of optimum (ideal) controllers in tasks that also are usually continuous, manual-control tasks. In an optimal control model, the simulated controller observes representations of a system's state variables (corrupted by sensory-system noise) and generates control responses (corrupted by motor-system noise) that minimize various error and cost criteria (for a review, see reference 9).

In contrast, network models developed in the Siegel-Wolf tradition usually represent operator tasks as organized sets of discrete subtasks. Typically, the representation of a complex task comprises a description of each of its subtasks and their organization. This description usually includes: (1) the conditions that must obtain before the subtask can begin, (2) the conditions obtaining at the end of the subtask, (3) the expected duration of the subtask (and the variability of its duration), and (4) the probability of successfully completing the subtask.

Control-theoretic and optimal-control models lend themselves most naturally to the description of continuous tasks. Their application, however, has not been limited to continuous tasks. An example is the Procedure-Oriented Crew Model (PROCRU, reference 10). The PROCRU model originated as a control-theory based model of the approach-to-landing stage of flight in a commercial airliner. It contains submodels describing flight control, display monitoring, communicating with air traffic controllers, and other flight activities. Similarly, although network models lend themselves most naturally to discrete tasks, their application has not been limited to discrete tasks. An example is the network model of the LHX helicopter developed in MicroSAINT by Laughery, Drews, Archer, and Kramme (11). One of the outputs of this model is a continuous variable whose value is an estimate of instantaneous operator workload during the course of a mission.

Because the psychometric models developed under this project follow a common plan and are written in a standard language, they are substantially easier to use than most computational performance models. Simulations can be specified, run, and analyzed using MicroSAINT's standard collection of menu-driven utilities. Thus, variables can be altered at the MicroSAINT Simulation Scenario menu. Data to be saved can be specified at the MicroSAINT Snapshots of Execution menu. Simulations can be run from the MicroSAINT Model Execution menu. Finally, data can be analyzed from the MicroSAINT Analysis of Results menu.

2. METHODS

The performance assessment test models we have developed follow the plan of the UTCPAB Generic Task. The Generic Task is a general model of the temporal organization of most of the tests of the UTCPAB. It also is as the basic plan followed by the computer programs of the UTCPAB Authoring System--the set of computer routines that make up the tests of the UTCPAB. Thus the models have the same temporal structure as the tests themselves. They represent the trial-by-trial temporal organization of behavior in the tests--the tests' performance structures.

EMPIRICAL PERFORMANCE DATA

We obtained estimates of the models' human-performance parameters from data provide by D. L. Reeves of the Naval Aercspace Medical Research Laboratory. The subjects were 28 male Naval and Marine Aviation Candidates. The data were obtained in a session comprised of four repetitions of a battery of tests drawn from the Walter Reed Performance Assessment Battery (WRAIRPAB; 12). An examination of the data indicated no significant change in the subjects' average performance across these sessions, so we derived our parameter estimates from all four repetitions of the tests. In general, the subjects' responses were sorted by correctness and reaction time (RT). The data were used to estimate the overall proportions of correct and incorrect responses (P(c) and P(i), respectively), the average correct- and incorrect-response RTs (RT(c) and RT(i), respectively) and the standard deviations of correct and incorrect single-trial RTs (SD(rt(c)) and SD(rt(i)), respectively).

SIMULATION PROCEDURES

MicroSAINT supplies gamma, normal, uniform, exponential, and Poisson random number generators. Of these, the exponential, Poisson, and gamma are skewed like empirical RT distributions. The exponential distribution, which is a special case of the gamma, yields only crude approximations to the shapes of empirical RT distributions. Both the Poisson and gamma resemble RT distributions qualitatively. The gamma distribution, however, applies more naturally than the Poisson to temporal variables (13). (The Poisson describes counts of exponentially-distributed variables.) The gamma also has two parameters v. the Poisson's one, which sometimes makes the gamma easier to fit. Based on these considerations, we used MicroSAINT's gamma-distributed random number generator to simulate RT in most of our models. This decision was made for the purpose of accurately describing the empirical data. We do not mean to suggest that gamma-distributed RTs necessarily follow from a theory of mental arithmetic (indeed, the data suggest otherwise).

Sequences of correct and incorrect responses, were simulated by treating responses as Bernoulli trials. Thus, the models generate RTs by drawing from simulated correct-response RT distributions on a randomly-determined 100(p(c))% of all trials. Similarly, the models draw RTs from a simulated, incorrect-response RT distribution on a random 100(1-p(c))% of all trials. The first-approximation models draw correct-response RTs from one probability distribution with a mean of RT(c) and a standard deviation of SD(rt(c)), and draws incorrect-response RTs from a second probability distribution with a mean of RT(i) and SD(rt(i)). (We will see, presently, that this strategy does not always work.)

3. RESULTS AND DISCUSSION

Figure 1 contains an example data set comprised of the overall RT histograms for correct and incorrect responses in the Serial Addition and Subtraction (SAS) test of the WRAIRPAB. (A fuller treatment of the data can be found in reference 14). The RT histograms were collapsed across subjects, test repetitions, and trials within repetitions. Several properties of the SAS RT distributions should be noted. First, the histograms have the positively skewed appearance typical of most RT distributions (13). Second, correct responses occur more frequently than errors (p(i) = 0.02 v. p(c) = 0.98). Third, correct-response reaction times are shorter, on average, than incorrect-response reaction times (RT(c) = 876.94 ms v. RT(i) = 1532.34 ms). Fourth, the variability of the correct-response reaction times is less than

the variability of the incorrect-response reaction times ($SD(rt(c))$ = 632.21 v. $SD(rt(i))$ = 1153.27). These are all standard results.

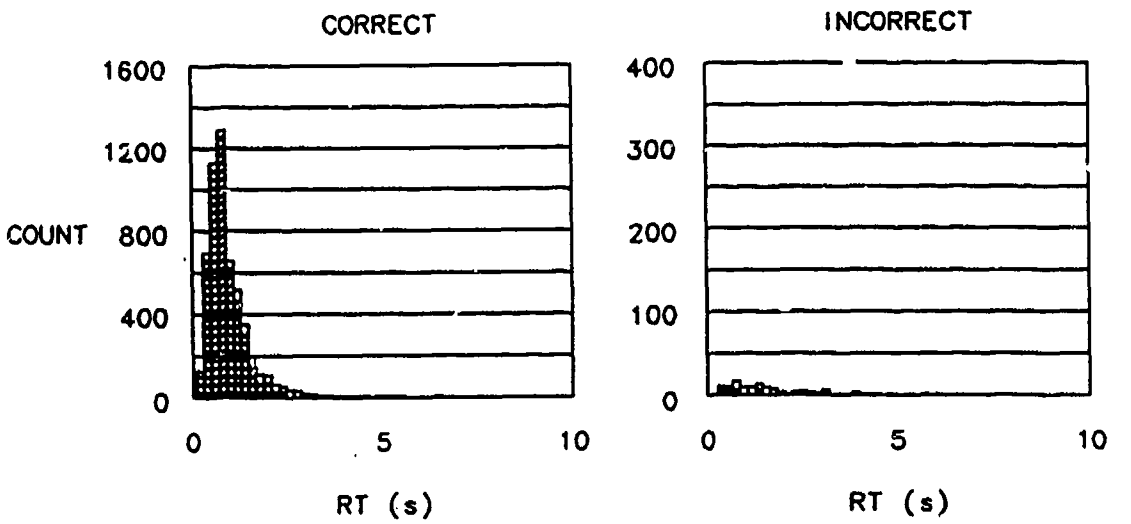

Figure 1. Correct- and incorrect-response reaction time (RT) distributions in Serial Addition and Subtraction.

Figure 2 illustrates observed and predicted correct-response RT distributions in Serial Addition and Subtraction. The function labeled 'Observed' is the empirical correct-response RT distribution. The function labeled "Full Data Set" is a gamma distribution with parameters (mean and variance) equalling the mean and variance of the empirical RT distribution. The correspondence is not especially close: the distribution of empirical RTs is much more peaked than the corresponding gamma distribution. A goodness-of-fit test using intervals containing expected frequencies of 5 or more yielded a Chi-square of 772.95 (df = 16, p < 0.005), which clearly allows us to reject the hypothesis that the empirical RTs arose from a gamma distribution with the same mean and variance as the data.

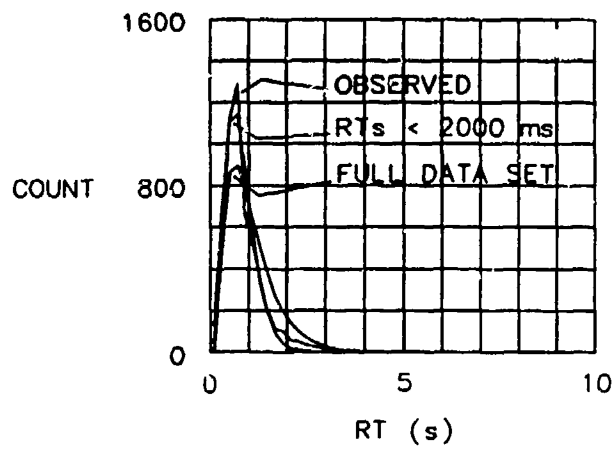

Figure 2. Observed and predicted correct-response reaction time (RT) distributions for Serial Addition and subtraction.

If the data in the tail of the empirical RT distribution are ignored the fit of the gamma to the empirical distribution is visibly improved. This is illustrated by the function labeled 'RTs < 2000 ms,' which is the gamma distribution with the same mean and variance as the subset of correct responses with RTs less than 2000 ms. The mean and variance of this distribution are 79? 235 ms and 143587 ms^2, respectively. A test of this distribution's goodness of fit also fails. The failure is somewhat less spectacular than before. A test calculated using the intervals with expected frequencies

of 5 or more yields a Chi-square of 597.21 (df = 9, $p < 0.005$). In this case, most of the discrepancy can be attributed to RTs in range of 1.5-2.5 s. In this region, the ordinate of the predicted curve falls well below that of the observed distribution (see the figure). Despite this result, the observed and predicted RT counts (in the intervals with more than 5 expected RTs) yield a highly respectable correlation (r = 0.9764).

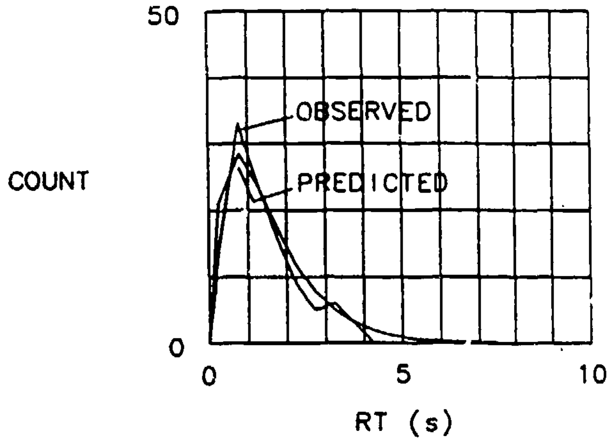

Figure 3. ***Observed and predicted incorrect-response reaction time (RT) distributions for Serial Addition and Subtraction.***

Figure 3 contains the observed and predicted incorrect-response RT distributions. In this case, a gamma distribution with mean and variance equal to those of the empirical incorrect-response RT approximates the empirical RT distribution well. The goodness-of-fit calculation, again based on intervals with more than 5 expected RTs, yields a Chi-square of 4.94 (df = 3, $p < 0.25$). The correlation between the observed and predicted RT-counts in those intervals again is quite high (r = 0.9473).

4. CONCLUSIONS

Models are abstract representations of systems. A model of a system consists of a set of important system variables and a set of relations among them. Models can be useful because they are compact relative to the systems they describe, and because they can be used to predict some of the effects of variation in system variables. A map, for example, is useful because it is more compact than the geography it describes and because it can be used to predict some of the consequences of changes in latitude and longitude.

The models we have described here and elsewhere are sequential-network designs; they are essentially task-analytic in nature. We think that modeling operational tasks in this fashion clearly represents an improvement in the quantitative description of human performance in operational systems. We also submit that computational models can also improve the quantitative description of performance in laboratory tests. This is partly because it is possible to develop models that retain the statistical properties of behavior that summary measures discard. Our models could, in fact, be expressed as equations. In part, this is because we have approximated the empirical performance data with probability distributions whose algebraic properties are well understood. We selected these distributions for reasons of computational efficiency. The penalty incurred was a loss of accuracy. Greater accuracy could be achieved, for example, by smoothing the empirical reaction-time histograms and sampling from the distributions thereby produced. Such nonparametric approaches to building models often produce results that are difficult or impossible to derive mathematically. Models based on theoretical considerations that are not easily related to well-developed bodies of statistical theory encounter similar problems (consider, for example, the difficulty of predicting the performance of neural networks). In such cases, computational procedures are often the only practical means of examining a problem.

An important question is whether a laboratory test that differs substantially from an operational behavior of interest can ever yield accurate predictions of real-life behavior. For example, to demonstrate that a

stressor affects human performance in an operational system requires one to show that the stressor changes the normal pattern of relations among the system, its operator, and the environment. A fairly direct approach to performing such a demonstration involves examining the stressor's effects in a hardware simulator (a flight simulator, for example). Simulator research, however, is slow and costly. Abstract, laboratory tests are faster, more economical. If properly carried out, laboratory tests should also produce more reliable results because more observations can be obtained at the same cost. Laboratory tasks, however, do not look like operational tasks. Consequently, they are often regarded with suspicion.

The only way to demonstrate empirically that performance on an abstract test predicts a variable's effects on performance in an operational system is to: (1) measure the effect of the variable on the test, (2) measure the effect of the variable on system performance, and then (3) show that these effects covary. However, this necessarily more than simply measuring the effect of the variable on operational performance. Thus, to justify the economics of such an enterprise, one must be able to say that any association found is reasonably likely to generalize to new tasks or new forms of operational performance. Asserting that a result will generalize, however, requires a separate appeal to theory or to a body of empirical evidence.

In principle, computational procedures can be used to amplify the information derived from the type of study just described. In particular, these techniques are useful for deriving predictions for new scenarios. This is exactly like deriving new predictions from theory. The process of deriving implications and then confirming or disconfirming them empirically is the pattern followed in the development of any body of scientific theory. Because computational techniques can accelerate the process of deriving predictions, they can improve the efficiency of experimentation: A well-designed simulation can rapidly explore the variable space of a theory for regions where its predictions are clearest. With this information, experiments can be optimized to provide strong tests of the theory by concentrating observations where they will do the most good. In this way, computational procedures can increase the rate at which useful information is acquired and, thereby, increase the range of phenomena that can be explored in a given amount of time.

REFERENCES

1. Hegge, F. W. "The JWGD3 MILPERF: A Coordinated and Integrated Tri-Service Program for Human Performance Assessment." In: Proceedings of the 1989 Medical Defense Bioscience Review, Columbia, MD. August, 1989, pp. 755-759.

2. Perez, W.A., Masline, P.J., Ramsey, E.G., and Urban, K.E., Unified Tri-services Cognitive Performance Assessment Battery: Review and Methodology, AAMRL-TR-87-007, Armstrong Aerospace Medical Research Laboratory, Wright-Patterson AFB, OH, March, 1987.

3. Englund, C.E., Reeves, D.L., Shingledecker, C.A., Thorne, D., Wilson, K.P, and Hegge, F.W., Unified Tri-service Cognitive Performance Assessment Battery (UTC-PAB): I. Design and Specification of the Battery, NHRC Report No. 87-10, Naval Health Research Center, San Diego, CA, 1987.

Shamma, S., Stanny, R., Laughery, R., Platt, C., & Sherry, D., Computer Aided Modeling of Cognitive Performance Assessment Tests Using the Micro SAINT Software, Institute for Statistical and Mathematical Modeling Technical Report, University of West Florida, Pensacola, March, 1989.

5. Stanny, R., Shamma, S., Laughery, R., Platt, C., Crisman, R., & Sherry, D. "Modeling the Unified Tri-Service Cognitive Performance Assessment Battery." In Proceedings of the 1989 Medical Defense Bioscience Review, Columbia, MD, August, 1989, pp. 793-796.

6. Shamma, S. E., Molina, E. A., & Stanny, R. R., MicroSAINT Programs for Numerical Methods of Integration and Differentiation, NAMRL Monograph 39, Naval Aerospace Medical Research Laboratory, Pensacola, FL, September, 1989.

7. Wortman, D.B., Duket, S.D., Seifert, D.J., Hann, R.L., and Chubb, G.P. Simulation using SAINT: A User-oriented Instruction Manual, Pritsker and Associates, West Lafayette, IN, 1978.

8. Siegel, A.I., and Wolf, J.J., Man-Machine Simulation Models, John Wiley and Sons, New York, NY, 1969.

9. Linton, P.M., Plamondon, B.D., Dick, A.O., Bittner, A.C., and Christ, R.E. "Operator Workload for Military System Acquisition." In: G.R. McMillan, D. Beevis, E. Salas, M.H. Strub, R. Sutton, and L. van Breda (Eds.), Applications of Human Performance Models to System Design, Plenum Press, New York, NY, 1989, pp. 21-45.

10. Baron, S., and Corker, K. "Engineering-Based Approaches to Human Performance Modeling." In: G.R. McMillan, D. Beevis, E. Salas, M.H. Strub, R. Sutton, and L. van Breda (Eds.), Applications of Human Performance Models to System Design, Plenum Press, New York, NY, 1989, pp. 203-217.

11. Laughery, K.R., Drews, C., Archer, R., and Kramme, K. "A MicroSAINT Simulation Analyzing Operator Workload in a Future Attack Helicopter." In: Proceedings of the National Aerospace and Electronics Conference, IEEE, New York, 1986.

12. Thorne, D.R., Genser, S.G., Sing, H.C., and Hegge, F.W., "The Walter Reed Performance Assessment Battery." Neurobehavioral Toxicology and Teratology, Vol. 7, pp. 415-418, 1985.

13. Luce, R.D., Response Times: Their Role in Inferring Elementary Mental Organization, Oxford University Press, New York, 1986.

14. Stanny, R.R., & Shamma, S.E. Models of Cognitive Performance Tests, Naval Aerospace Medical Research Laboratory, Pensacola, FL [Submitted].

BIBLIOGRAPHY: FY88-FY89

Shamma, S., Stanny, R., Laughery, R., Platt, C., & Sherry, D., Computer Aided Modeling of Cognitive Performance Assessment Tests Using the Micro SAINT Software, Institute for Statistical and Mathematical Modeling Technical Report, University of West Florida, Pensacola, March, 1989.

Stanny, R., Shamma, S., Laughery, R., Platt, C., Crisman, R., & Sherry, D. "Modeling the Unified Tri-Service Cognitive Performance Assessment Battery." In Proceedings of the 1989 Medical Defense Bioscience Review, Columbia, MD, August, 1989, pp. 793-796.

Shamma, S. E., Molina, E. A., & Stanny, R. R., MicroSAINT Programs for Numerical Methods of Integration and Differentiation, NAMRL Monograph 39, Naval Aerospace Medical Research Laboratory, Pensacola, FL, September, 1989.

INDIVIDUALS RECEIVING PAY AND/OR OVERHEAD COST SUPPORT: FY88-FY89

LCDR R. P. Crisman, NAMRL
Dr. A. Koblasz, Georgia Technical Institute (IPA)
Dr. W. A. Morey, NAMRL
Mr. D. Prettyman, NAMRL
Dr. S. E. Shamma, University of West Florida (IPA)
Ms. I. K. Vogel, NAMRL